AFRL-SR-BL-TR-98-

0299

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

FINAL TECHNICAL REPORT

April 1, 1995 - December 31, 1997

Grant No. F49620-95-1-0265

"VISUAL PROCESSING OF OBJECT VELOCITY AND ACCELERATION"

Principal Investigator: Suzanne P. McKee, Ph.D.

Institution: The Smith-Kettlewell Eye Research Institute
2232 Webster Street
San Francisco, CA 94115

REPORT DOCUMENTATION PAGE

Form Approved

OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 27, 1998	FINAL TECHNICAL REPORT 04/01/95-12/31/97

4. TITLE AND SUBTITLE

"VISUAL PROCESSING OF OBJECT VELOCITY AND ACCELERATION"

5. FUNDING NUMBERS

G - #F49620-95-1-0265
TA# - 2313/BS

6. AUTHOR(S)

Suzanne P. McKee, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Smith-Kettlewell Eye Research Institute
2232 Webster Street
San Francisco, CA 94115

8. PERFORMING ORGANIZATION REPORT NUMBER

CFDA #12.800

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NL
110 Duncan Avenue Room B115
Bolling AFB DC 20332-8050

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by the documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This research analyzed human detection of object motion in random motion noise. Results from experiments demonstrated that 1) a flexible neural network enhances detection of a feature moving in a constant direction or changing direction slowly; 2) the network uses a highly non-linear facilitation, rather than the linear sum of contrast or luminance signals; 3) a stimulus pattern that moves along the motion path is more detectable than one oriented orthogonal to it; 4) stereopsis does not provide much benefit for detecting motion signals in noise; 5) motion along the z-axis (motion-in-depth) can be masked by static disparity noise; 6) the motion system can simultaneously encode both the local directions of small features and the global direction of the flow field. A Bayesian model of object motion and surface segmentation is being developed to explain these observations. Future work will explore whether this model of human motion processing can be implemented computationally in VLSI hardware for detecting moving projectiles in the midst of noise. Twelve papers were accepted for publication in refereed journals; four chapters were also supported by this grant.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Trajectory signal, Motion signal, Motion and Stereopsis	10
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500 **Standard Form 298 (Rev.2-89)** Prescribed by ANSI Std. 239-18 298-102

TECHNICAL SUMMARY OF WORK ACCOMPLISHED

1. The Trajectory Network

In our first paper on trajectory detection (Watamaniuk, McKee & Grzywacz, 1995), we showed that a single dot moving in a consistent direction is highly visible when presented at a randomly-chosen location in the midst of identical dots moving in Brownian motion (random direction noise). We demonstrated that detection was based on signal motion, rather than on the pattern traced by the trajectory, i.e., a 'string' of collinear dots. All of the dots in the display were actually 'hopping' in apparent motion, so the temporal order of the steps that defined the trajectory signal could be randomized, preserving the collinear pattern, but not the motion sequence. When we randomized the temporal sequence, signal detection fell to chance, establishing that motion was necessary for detection. We also found that the signal did not have to move on a straight trajectory to be easily detected. A signal dot that changed direction gradually was also quite visible; in fact, an extended circular trajectory, which rotated through 60 deg/100 msec, was almost as detectable as a straight trajectory.

In the second paper (Grzywacz, Watamaniuk & McKee, 1995), we presented a modified version of "Temporal Coherence Theory" (Yuille & Grzywacz, 1988; Grzywacz, Smith, & Yuille, 1989) that explained our results quantitatively. In this model, detection of the trajectory signal was enhanced by a flexible network that connected motion units of the same spatial scale. Signals from a currently stimulated motion unit were fed forward to units tuned to the same or similar directions, increasing their responsiveness to subsequent stimulation by the moving trajectory signal. This facilitation, which implements a kind of temporal smoothing, was in competition with another process that degraded signal detection. The competing process in the model was a second smoothing operation, called spatial coherence, that minimized local differences among similarly-tuned units within a particular spatial neighborhood. If the trajectory dot encountered a noise dot moving in the same direction, the signal generated by the trajectory was reduced substantially by spatial coherence. Our simulations demonstrated that this model could predict the psychophysical data in detail.

Recently, Dr. Preeti Verghese joined Smith-Kettlewell, replacing Dr. Scott Watamaniuk on this project. She was not convinced that our results ruled out an explanation based on the signals generated by local motion units, acting independently. To be highly detectable, the trajectory signal had to be presented for a duration of 200 - 400 msec. Perhaps the combined responses from several independent local units all stimulated by the trajectory were sufficient to explain human performance without the addition of a facilitatory net. Even the detection of circular trajectories might be explained by a motion unit that responded to the oblique direction defined by the implicit chord across the circular path. She noted that an extended (>200 msec) trajectory was detected on *all* trials, if the observer knew where it would be presented in the noise display -- a result that suggested

that a relationship to the abundant literature on visual search (Treisman & Gelade, 1980). Rather than the more elaborate theoretical structure proposed in the Temporal Coherence paper, Dr. Verghese asked whether the austere 'ideal observer' models that had been constructed to explain performance in visual search (Palmer, Ames, & Lindsey, 1993; Verghese & Nakayama, 1994; Geisler & Chou, 1995) might be sufficient to explain our results.

To explore this question, Dr. Verghese began by calculating the response of a "Motion Energy" unit (Adelson & Bergen; 1985)[1] to an extended trajectory in our noise conditions. To estimate the optimum space constant for this model unit, she computed the response of units of varying size to presentations of signal plus noise, as well as to presentations of noise alone. The unit size that yielded the highest signal-to-noise ratio was deemed optimal. Her calculations showed that the optimal unit would respond to about seven frames of the trajectory (~100 msec), thereby confirming an earlier calculation by Dr. Grzywacz based on a different statistical criterion (Grzywacz, Watamaniuk & McKee, 1995). Larger units with longer time constants have poorer signal/noise ratios, because while they see more of the extended trajectory, they also see more random noise[2].

We next showed that this model unit could explain human performance under specified experimental conditions. We presented a 100 msec (7 frame) trajectory within a small aperture in dense noise; the aperture was just large enough to permit unobstructed viewing of a straight trajectory moving in one of eight directions chosen at random. Human observers judged in which of two intervals the trajectory was present, and which contained noise alone (standard 2IFC paradigm). For this simple judgment, observers averaged 81% correct. To calculate model unit performance, Dr. Verghese convolved the unit's spatial-temporal response function with the two-intervals of each trial, and choose the interval with the larger response (max rule), scoring the unit's performance just like the human performance. These simulations showed that such a unit would give a larger response on 83% of trials to the 100 msec trajectory signal plus noise, than to the noise alone. The model units are ideal detectors and have no added noise, so the good agreement between the simulated response and human performance indicates that the stimulus noise is substantially larger than the internal noise of the human observer. The agreement between data and model also indicates that the spatio-temporal constants chosen for the model units are plausible. Note that, for this simulation, the trajectory signal was in known location, i.e. within a small specified aperture.

If the 100 msec trajectory signal is presented anywhere at random within a large noise field, then an 'ideal observer' must monitor all the local motion units which tile the

[1] The elaborated Reichardt correlator proposed by van Santen & Sperling (1985) would produce similar results for the conditions described here.

[2] The motion units were assumed to be roughly circular, consistent with much psychophysical data (Anderson & Burr, 1985; Anderson *et al*, 1991).

noise field. Under these circumstances, the model performance for detecting a single 100 msec trajectory segment at an unknown location plummeted to about 58% correct. Model performance is poor because there is a high probability that one of the many local motion units responding to a noise location in the 'noise' interval will have a larger response than a unit responding to either the signal or a noise location in the 'signal' interval, on a large number of trials. Any factor which increases the number of units that this 'ideal observer' has to monitor, e.g., an increase in the size of the noise field, will degrade performance, because effectively the number of noise samples is increasing without a commensurate increase in the number of signal samples.

What about the response of the motion units to an extended trajectory? A 200 msec trajectory, presented at a randomly-chosen location within a 4 x 4 deg square region centered in the midst of dense noise, is correctly detected by human observers on ~80% of trials. Since one 100 msec trajectory segment fits the spatio-temporal constraints of our model unit, the simplest assumption is that the 200 msec trajectory stimulates two non-overlapping units with independent signal samples. Are two signal-driven units sufficient to explain human performance? Before calculating the response of our model units, we first asked how many 100 msec trajectory segments presented at random locations within the central square region (Figure 1A) are needed to produce human detection equivalent to the detection of the 200 msec trajectory. Figure 1B shows that 5 - 7 100 msec segments are needed to equal the detection of the longer trajectory (hatched region at top of figure). In short, a 200 msec trajectory is much more than the sum of its parts. The dotted line passing through the human data shows the prediction of the independent local units model. The model does an excellent job of describing human performance for the short motion segments scattered throughout the noise, but fails for trajectory detection, since the prediction for two independent units is well below human detection of the 200 msec trajectory.

Perhaps the 200 msec trajectory stimulates more than two local units. What about the overlap between local units responsive to the trajectory? We calculated signal improvement if we assumed massive overlap of units responding to the 200 msec trajectory moving in a known direction at a known location. Predicted detection increased to 74%, but still short of human performance. The problem with increasing the units' overlap for the trajectory is that, for the general case of a 200 msec trajectory moving in a random direction in an unknown location, a similar overlap would have to be assumed for units tiling the noise field, thereby substantially increasing the number of noise samples. We have not performed this calculation, but we suspect that the increase in the number of signal samples would be more than offset by the increased number of noise samples. Would different types of motion units and/or different decision rules explain human detection of extended trajectories? Possibly, but our work on stimulus configuration provides additional evidence for the existence of a facilitatory network in human motion processing.

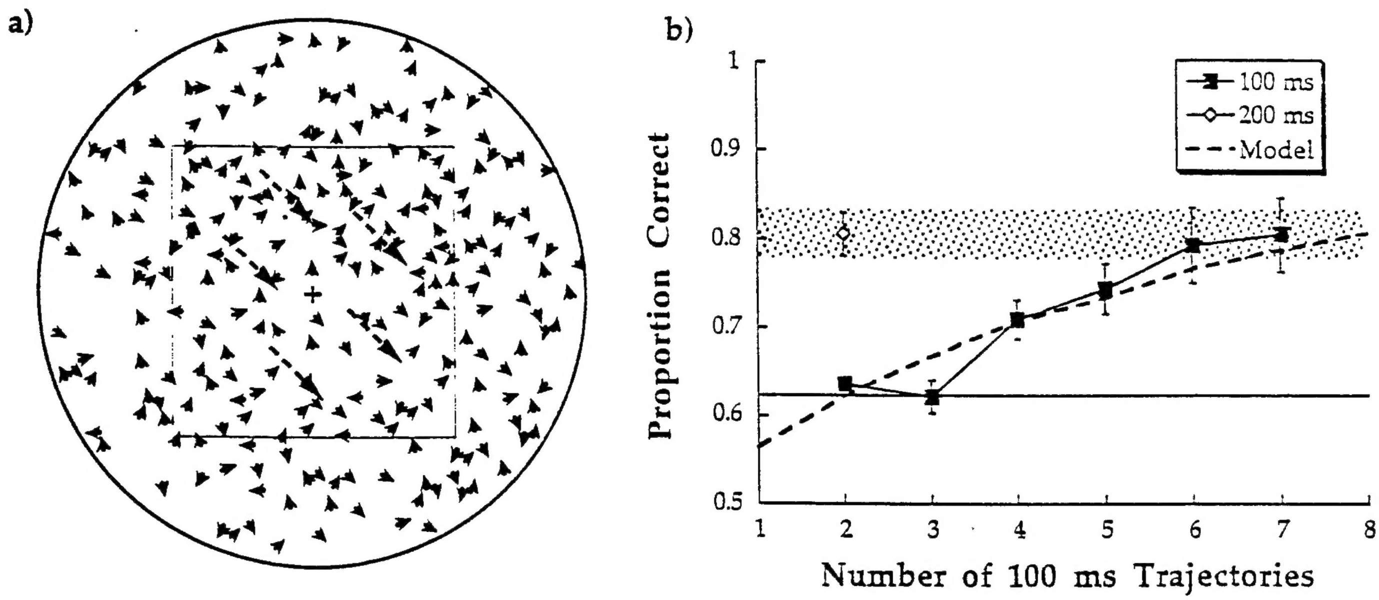

Figure 1

We measured trajectory detection for rigid triplets of dots, arranged either perpendicular to the direction of motion or parallel to it, as a function of the separation between the aligned dots (see Figure 2). As in our earlier work, the triplet trajectories were presented at a random location in the central region of the noise field. For all tested separations, the parallel configuration was more easily detected. The difference between the two configurations was not due to temporal summation of contrast or luminance signals generated by consecutive dots in the parallel configuration. We measured contrast thresholds as a function of the time between the presentation of two (or three) dots at the same location, in a conventional temporal summation paradigm (Watson, 1986). The temporal summation functions indicated that the benefits from spatial coincidence of the dots last for approximately 50 msec whereas the increased detectability of the parallel configuration is observed up to the largest separations tested. At the largest separation (2.5 deg) and a speed of 12 deg/sec, a trailing dot will reach the same position as a lead dot ~210 msec later.

We considered whether an alternative to the circular receptive field of the 'motion energy' units might explain these results, e.g. a motion unit with a receptive field that is spatially-elongated (Fredericksen, Verstraten & van de Grind, 1994). Our results indicate that the temporal sequence of the motion segments in an extended trajectory matters as much as their spatial arrangement. If a 200 msec trajectory is divided into two 100 msec segments, and the second 100 msec segment is presented spatially in front of the first 100 msec segment, detection is weaker than for the sequential order appropriate to natural motion (see Figure 3). Therefore, a unit that accounts for our results would have to be elongated in space and time. While this type of receptive field would explain our triplet configuration findings, it cannot explain the high detectability of circular trajectories.

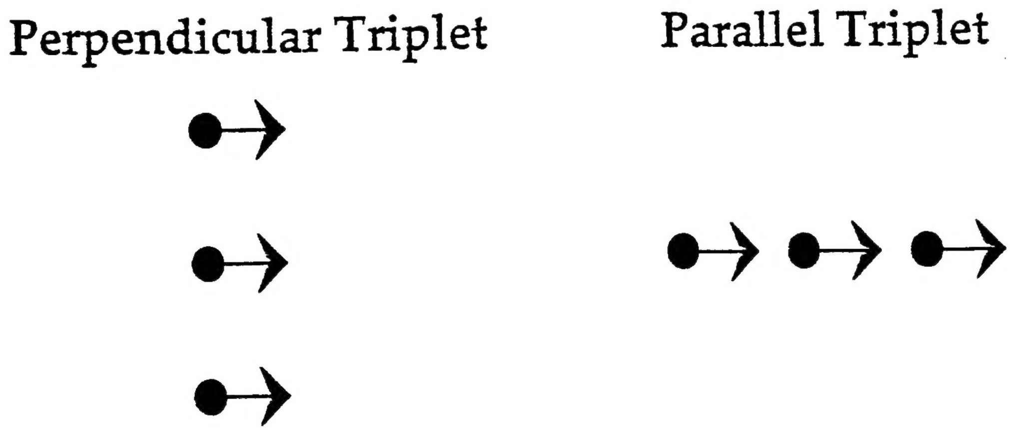

Figure 2

Therefore, we speculate that the enhanced detection of the parallel triplet is due to the *forward* propagation of a facilitatory signal, which subsequent dots catch up with. If so, the facilitatory signal propagated along the network has a long decay time.

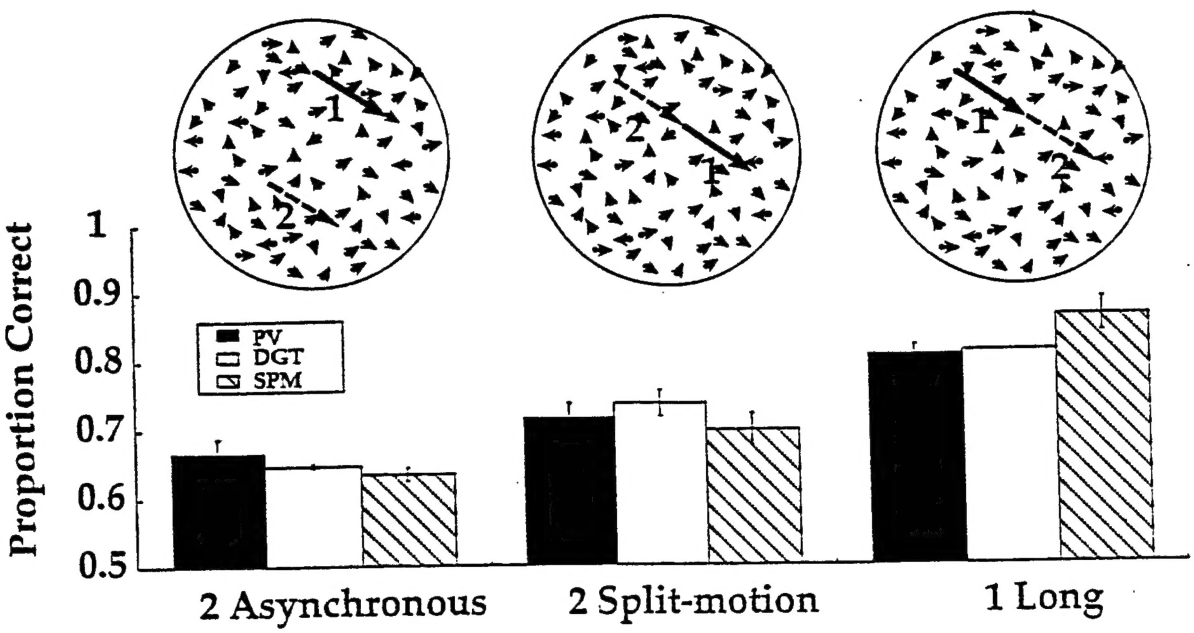

Figure 3

2. Motion and Stereopsis

Our laboratory has been exploring the relationship between human motion processing and human stereopsis. Stereopsis is a sluggish system, which takes considerable time to reach its best sensitivity (Ogle & Weil, 1958; McKee, Levi & Bowne, 1990), while motion processing must necessarily operate on a rapid time course if it is to deliver information which can be used to guide human movements. Thus, poor speed discrimination for laterally-moving targets, defined only by stereopsis (no coherent motion

in the monocular half-images) might seem fairly predictable. Harris & Watamaniuk (1996) found that Weber fractions for speed discrimination were over 0.3 for cyclopean targets, much inferior to the typical Weber fractions for luminance-defined targets which were between 0.05 -0.1 for comparable conditions. Patterson, Donnelly, Phinney, Nawrot, Whiting & Eyle (1997) reported similar poor speed discrimination for cyclopean targets.

McKee, Watamaniuk, Harris, Smallman & Taylor (1997) have recently demonstrated the poor disparity tuning of motion units. As shown by the diagram in the left half of Figure 4, motion noise was presented in two planes that straddled the fixation plane where the trajectory signal was presented. By comparing trajectory detection to a similar detection task for static targets, McKee *et al* found that motion units were far less sensitive to disparity than units responding to static targets (see graphs on the right of Figure 4).

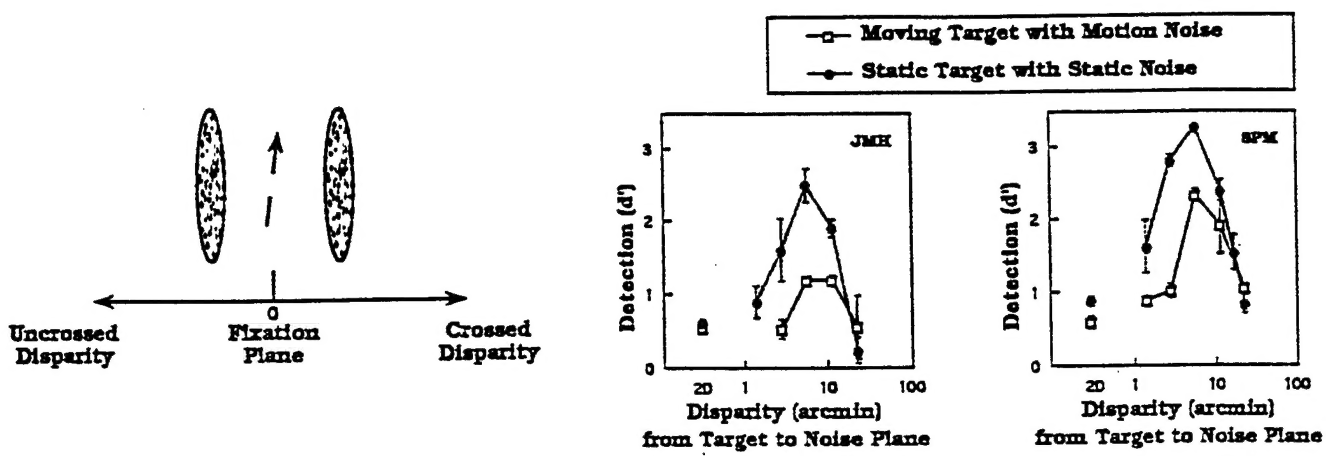

Figure 4

What about motion-in-depth? Regan and colleagues have long argued for a special mechanism that responds to motion along the z-axis, separate from the mechanisms that respond to static position disparity or to lateral motion. As strong support for this premise, they found subjects who had normal sensitivity to lateral motion and static disparity, but were unable to detect motion in depth at specified loci in the visual field (Regan, Erkelens, Collewijn, 1986; Hong & Regan, 1989). However, our results suggest that this special mechanism may not exist at fine scales in the central fovea.

Using luminance-defined targets (bright points), Harris, McKee, & Watamaniuk (1997) examined the masking effects of static disparity noise on the detection of motion-in-depth. Subjects were asked to detect a single point moving along the z-axis, presented at an unknown

location in the midst of a three-dimensional cloud of static points. Detection of the z-axis motion was compared to detection of the lateral motion produced by viewing one stereo half-image of the same display (lateral speed = half speed of z-axis motion). Subjects easily detected both the z-axis motion and the lateral motion of the half-image when viewed with a single static reference point. However, the addition of static 3D noise profoundly masked detection of motion-in-depth (Figure 5). As few as eight static noise points reduced motion-in-depth detection from over 90% correct to about 65% correct. Neither two-dimensional nor three-dimensional static noise had much effect on the detection of the slow lateral motion associated with the half-image. Our results show that motion-in-depth at fine scales in the fovea is mediated by some mechanism sensitive to static disparity noise. The most likely explanation is that this foveal mechanism is simply composed of the primary disparity units of the stereo system.

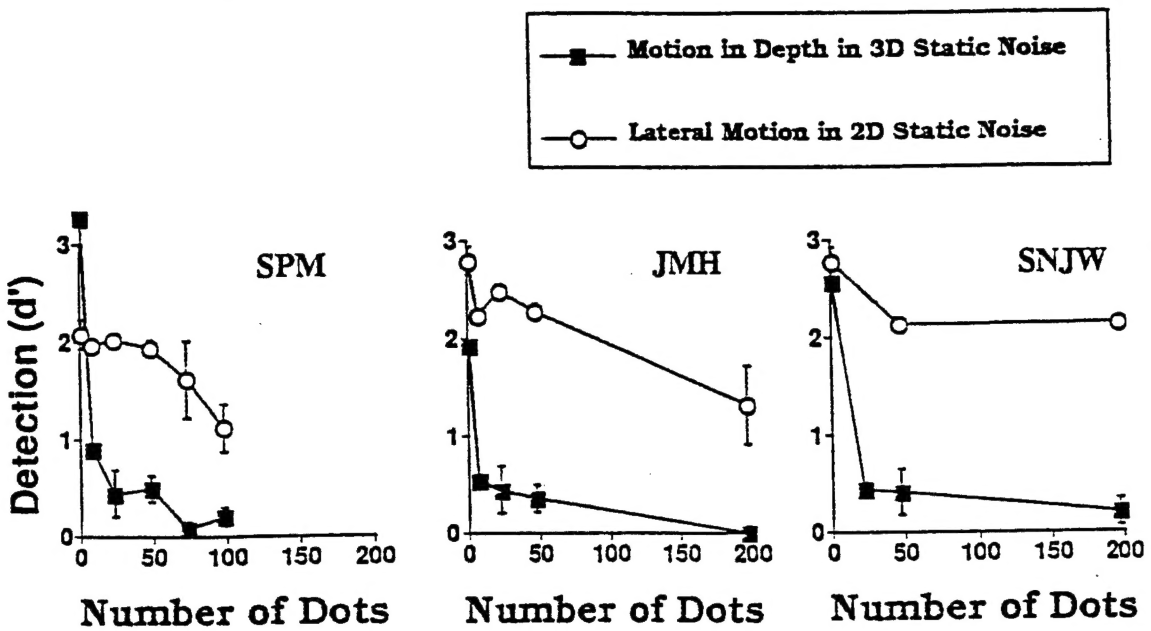

Figure 5

Despite the mismatch between the temporal characteristics of these two systems, there is considerable evidence for a synergistic relationship between stereopsis and motion processing in the some level in human visual system. Johnston, Cumming and Landy (1994) reported that motion parallax information is used to correct the errors introduced into shape estimation by the stereo system. Stereo information is likewise used to resolve ambiguities in motion displays. In the "Kinetic Depth Effect", two-dimensional line figures rotated around a vertical axis can appear to be rigid three-dimensional shapes, but the direction of rotation and the depth order (front and back) are ambiguous (Wallach and O'Connell, 1953). For many observers, the addition of unambiguous stereo information about the depth order resolves the rotational ambiguity (Dosher, Sperling & Wurst, 1986). A similar disambiguating effect of stereopsis has been observed for rotating transparent

cylinders composed of moving random dots, where again the direction of rotation and depth order are ambiguous (Nawrot & Blake, 1991). When a pair of one-dimensional moving gratings at different orientations are superimposed, they can cohere into single pattern with a unique direction ("plaid"), or drift across one another like transparent surfaces. Whether the plaid percept is coherent or transparent is affected by the relative disparity of the component gratings (Trueswell & Hayhoe, 1992). Shimojo, Silverman & Nakayama (1989) found that the influence of 'terminators' in the barber-pole illusion was reduced if the occluding aperture was presented at a crossed disparity in front of the moving lines.

We can reconcile these diverse findings by assuming that, after initial encoding by the primary units for each dimension independently, stereo information is combined with motion information at neural sites responsible for image segmentation and surface representation, as suggested by Alais, van der Smagt, Verstraten & van de Grind (1996). In this type of organization, disparity will not greatly influence motion detection, speed or direction discrimination, since those judgments depend either on motion signals generated in the primary motion units or on motion networks composed of such units. However, subsequent integration of independent information from both systems will permit stereo disambiguation of motion-defined surface structure.

3. Other Work

In the Welch, MacLeod & McKee (1997) study, temporal order discrimination for a pair of sequentially-presented points (which point came on first?) was used to probe the spatial and temporal constraints of the trajectory network. A single perturbing point presented before the onset of the test pair could strongly affect the perceived order, i.e., the direction of motion. The direction defined by the onset of the perturbing point and the first member of the test pair, nulled the directional signal defined by the test pair itself, if it was opposite to the perturbing direction. To overcome the perturbing effect, the time separating the test pair had to be increased drastically (5x). The nulling effect was maximal when the perturbing point preceded the test pair by 100 msec, but it extended for durations up to 200 msec. The largest effects were found when the spacing between the perturbing point and the test pair was equal to separation between the test pair, consistent with a scale-dependent interaction.

Pettet, McKee and Grzywacz (1997) applied a <u>static</u> spatial coherence model to the detection of contours formed of Gabor patches, presented in noise consisting of several hundred Gabor patches with random positions and orientations. This static model had much in common with the temporal coherence model described above for moving targets. In agreement with previous studies (Kovacs & Julesz, 1993), we found that closed contours were more easily detected than open contours. However, the introduction of two sharp changes in orientation (>30 deg) between neighboring Gabor patch elements in closed-path contours reduced detection performance to the same levels obtained with open-

ended contours. These psychophysical data were in accord with the results of model simulations. We concluded that closure alone is not sufficient to enhance the visibility of a contour. Rather, if a closed contour meets certain geometric constraints, then lateral interactions based on these constraints may generate facilitation that reverberates around the closed path, thereby enhancing the contour's visibility.

Publications

Papers:

Bravo, M. J. & Watamaniuk, S. N. J. (1995). Evidence for two speed signals: a coarse local signal for segregation and a precise global signal for discrimination. *Vision Research*, **35**, 1691-1698.

Watamaniuk, S.N.J., McKee, S.P. and Grzywacz, N. (1995) Detecting a trajectory embedded in random-direction motion noise. *Vision Research*,.**35**, 65-78.

Harris, J.M. and Watamaniuk, S.N.J. (1995) Speed discrimination of motion-in-depth using binocular cues. *Vision Research*, **35**, 885-896.

Grzywacz, N.M.,Watamaniuk, S.N.J. and McKee, S.P. (1995) Temporal coherence theory for the detection and measurement of visual motion. *Vision Research*, **35**, 3183-3203.

Watamaniuk, S.N.J. and McKee, S.P. (1995) Seeing motion behind occluders. *Nature*, **377**, 729-730.

Grzywacz, N.M. & Norcia, A. M. (1995) Directional selectivity in the cortex, in Arbib, M.A.,*The Handbook of Brain Theory and Neural Networks*. Cambridge,MA: MIT Press, 309-311

Harris, J.M. and Watamaniuk, S.N.J. (1996) Poor speed discrimination suggests that there is no specialized speed mechanism for cyclopean motion. *Vision Research*, **36**, 2149-2157.

McKee, S.P., Watamaniuk, S.N.J., Harris, J.M., Smallman, H.S. and Taylor, D.G.(1997) Is stereopsis effective in breaking camouflage for moving targets? *Vision Research*, **37**, 2047-2055.

Welch, L., MacLeod, D.I.A. and McKee, S.P.(1997), Motion interference: Perturbing perceived direction. *Vision Research*, **37**, 2725-2736.

Watamaniuk, S.N.J. and McKee, S.P. (1998), Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics*, **60**, 191-200.

Pettet, M.W., McKee, S.P., & Grzywacz, N.M. (1998) Constraints on long range interactions mediating contour detection. *Vision Research* , **38**, 263-274..

Yuille, A.L., Burgi, P.Y. & Grzywacz, N.M.(1997) Visual motion estimation and prediction : a probabilistic network model for temporal coherence. in *Proceedings of International Conference on Computer Vision*. Bombay, India In Press.

Verghese, P., Watamaniuk, S.N.J., McKee, S.P. & Grzywacz, N.M. (1997) Local motion detectors cannot account for the detectability of an extended trajectory in noise. *Vision Research*. In Press.

McKee, S.P. and Smallman, H.S.(1997), Size and Speed Constancy in *Perceptual Constancies: Why things look as they do*. Walsh, V. and Kulikowski, J.J. (Eds.) Cambridge University Press, New York, In Press

Vaina, L.M., Gryzwacz, N.M., LeVay, M., Bienfang, D. & Wolpow, E. (1997), Perception of motion discontinuities in patients with selective motion deficits. in Watanabe,T. *High-Level Motion Processing: Computational, Physiological and Psychophysical approaches*, North Holland Press, In Press.

Yuille, A.L. & Grzywacz, N.M. (1997) A theoretical framework for visual motion. in Watanabe,T. *High-Level Motion Processing: Computational, Physiological and Psychophysical approaches*, North Holland Press, In Press.

Abstracts

McKee, S.P., Harris, J.M. & Watamaniuk, S.N.J. (1995) Detection of motion in depth is disrupted by static disparity. (ECVP Abstracts) *Perception Supplement*, **24**, 33.

Harris, J.M & Watamaniuk. S.N.J. (1995) A poor speed signal for disparity-defined motion. *Investigative Ophthalmology & Visual Science*, **36**, S813.

Verghese, P. McKee, S.P., Watamaniuk. S.N.J. & Grzywacz, N.M. (1996) Evaluating a simple model for detecting signals in noise *Optical Society of America, October 1996 meeting*

Verghese, P., Watamaniuk, S.N.J., McKee, S.P. & Grzywacz, N.M.(1997) Stimulus configuration determines the detectability of motion signals in noise. *Investigative Ophthalmology & Visual Science*, **38**,

Harris, J.M, McKee, S.P. & Watamaniuk. S.N.J. (1997) Motion-in-depth and lateral motion are detected by different mechanisms. *Investigative Ophthalmology & Visual Science*, **38**,

McKee, S.P. (1997) Does stereopsis enhance motion processing? in *ATR Workshop on Human Visual Perception: Linking Psychophysics and Engineering*. 16-19.

Personnel Supported and/or Associated with Project:

Suzanne P. McKee, Ph.D. Principal Investigator

Norberto M. Grzywacz, Ph.D., Co-investigator

Alan L. Yuille, Ph.D., Co-Investigator

Preeti Verghese, Ph.D., Associate Scientist

Mark Pettet, Ph.D., Postdoctoral Fellow

Douglas G. Taylor, B.S., Programmer-Engineer

Scott N.J. Watamaniuk, Ph.D., Co-investigator (now Asst. Professor, Wright State University

Julie M. Harris, Ph.D., Postdoctoral Fellow (now Lecturer, University of Edinburgh)

Harvey S. Smallman, Ph.D., Postdoctoral Fellow (now Asst. Research Professor, UCSD)

Pierre-Yves Burgi, Ph.D., Associate Scientist